

# **ELABORACIÓN DE UNA CLASIFICACIÓN NO SUPERVISADA Y SUPERVISADA PARA GENERAR LAS COBERTURAS VEGETALES DE UNA IMAGEN SATELITAL LANDSAT 7 - ETM USANDO LOS PROGRAMAS R Y PCI GEOMATICS CON EL FIN DE COMPARAR LOS RESULTADOS OBTENIDOS**



**Trabajo de grado para optar por el título:  
Especialista en Geomática  
Realizado por:  
Andrea Patricia Acosta Ovalle  
Código: 3101372**

**UNIVERSIDAD MILITAR NUEVA GRANADA  
FACULTAD DE INGENIERÍA  
ESPECIALIZACION EN GEOMATICA  
BOGOTA D.C  
2017**

# **ELABORACIÓN DE UNA CLASIFICACIÓN NO SUPERVISADA Y SUPERVISADA PARA GENERAR LAS COBERTURAS VEGETALES DE UNA IMAGEN SATELITAL LANDSAT 7 - ETM USANDO LOS PROGRAMAS R Y PCI GEOMATICS CON EL FIN DE COMPARAR LOS RESULTADOS OBTENIDOS.**

**Elaboration of an unsupervised and supervised classification to generate the vegetal covers of a LANDSAT 7-ETM satellite image using the R and PCI geomatics programs in order to compare the results obtained.**

Andrea Patricia Acosta Ovalle  
Ingeniera Catastral y Geodesta  
Estudiante de Especialización en Geomática  
Bogotá, Colombia  
[andreaacostaovalle@gmail.com](mailto:andreaacostaovalle@gmail.com)

## **RESUMEN**

En la actualidad existen diversas alternativas para procesar imágenes satelitales, ya sea con software libre o comercial, ambas opciones son altamente competitivas y poseen gran cantidad de características de procesamiento. Para este estudio se seleccionaron PCI y R que permiten realizar una clasificación no supervisada y supervisada para comparar los resultados obtenidos y evaluar las ventajas y desventajas al usar dichos algoritmos de clasificación, calificando aspectos técnicos como exploración de estadísticas descriptivas preliminares, funcionabilidad, tiempos de ejecución, procesos y presentación de resultados.

**Palabras Clave** Clasificación, imágenes satelitales, coberturas vegetales, comparación.

## **ABSTRACT**

Nowadays several alternatives to process satellite images exist, either with free or commercial software, both options are highly competitive and present a large number of processing characteristics. For this study, PCI and R were selected to allow an unsupervised and supervised classification for comparing the results obtained and to evaluate the advantages and disadvantages when using those classification algorithms, by qualifying technical aspects such as exploration of preliminary descriptive statistics, functionality, execution times, processes and results presentation.

**Keywords:** Classification, satellite images, vegetation coverage, comparison.

## **INTRODUCCIÓN**

Las imágenes satelitales son de fácil acceso, tienen infinidad de aplicaciones y suministran información de una gran extensión de la superficie terrestre, convirtiéndose en un insumo necesario para la elaboración de proyectos geomáticos.

Hoy en día es de gran utilidad elaborar análisis en detección de cambios de patrones y de vegetación, ya que entre sus múltiples aplicaciones permite diferenciar los tipos de vegetación existente en una en un área determinada sin tener que realizar vistas a campo permanentemente [2], evalúa los recursos naturales existentes, mejora la toma de decisiones y se puede hacer un seguimiento constante a dichos procesos. Por tal razón, es importante desarrollar e implementar software que lleven a cabo análisis geoespaciales con alta precisión y exactitud.

Este estudio busca realizar una comparación entre los software R y PCI para obtener datos cualitativos y cuantitativos que permitan a partir de los componentes estadístico y geoespacial, identificar los principales cambios en la técnica aplicada para reconocer de forma preliminar las coberturas vegetales [4] presentes en la imagen Landsat 7 – ETM del 2003; como resultado final se generarán los mapas temáticos de las coberturas vegetales obtenidas con los programas de procesamiento seleccionados y se evaluarán las ventajas y desventajas al usar diferentes algoritmos de clasificación permitiendo que los usuarios puedan seleccionar la herramienta que más se ajusta a sus necesidades.

## **1. ANTENDENTES**

Las técnicas de geoprocesamiento nacen por la necesidad de conocer la superficie de la tierra y han evolucionado a través del tiempo, permitiendo obtener productos de alta calidad y disponibles para muchos campos científicos.

Las primeras plataformas de observación fueron globos aerostáticos dotados de cámaras que tomaban pocas fotografías, en un vuelo lento y cubrían un área de estudio pequeña. En los años siguientes, debido a conflictos bélicos se impulsa la observación fotográfica desde un avión y se diseñan nuevas técnicas de adquisición y procesamiento fotográfico.

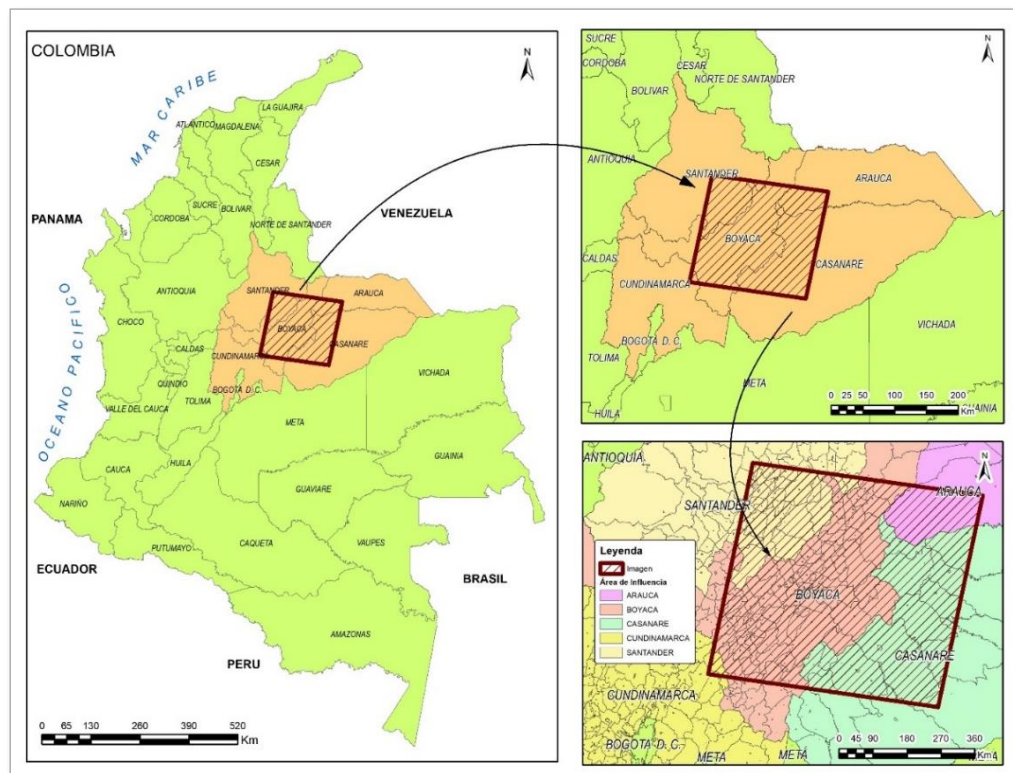
Gracias al nacimiento de la era espacial, aparece la idea de realizar observaciones sistemáticas de la tierra desde el espacio, enviando satélites meteorológicos con cámaras de baja resolución, que inicialmente solo permitían discriminar entre nubes, nieve y agua; los avances de estos satélites y las primeras misiones espaciales tripuladas adquirieron datos de los recursos naturales e impulsaron el desarrollo por la NASA (National Aeronautics and Space Administration) del programa ERTS (Earth Resources Technological Satellite) conocido también como LANDSAT, impulsando a muchos países a disponer de medios propios para cartografiar y evaluar los recursos naturales existentes.

Por todo lo anterior, se ha tenido que diseñar e implementar software capaz de extraer información y procesar las miles de imágenes generadas cada día para aplicaciones de uso militar y civil; estos programas se han dedicado principalmente a ofrecer alternativas para clasificar las coberturas del suelo y de ahí derivar otros estudios.

## **2. METODOLOGÍA**

### **2.1 DESCRIPCIÓN ZONA DE ESTUDIO**

Para el desarrollo de este proyecto se seleccionó como insumo principal una imagen satelital Landsat 7 - ETM del año 2003 ubicada de acuerdo a la grilla mundial en un Path (franja vertical) de 007 y un Row (fila horizontal) de 056 en el territorio colombiano que comprende 5 departamentos, como se muestra en la Fig. 1.



**Fig. 1.** Localización general zona de estudio.  
**Fuente:** Elaboración Propia a partir de la Información base del IGAC 2014.

## 2.1.1 Límites geográficos

**Tabla 1 - Información General**

<b>Cundinamarca</b>	
<b>Superficie</b>	24.210 Km <sup>2</sup>
<b>Gentilicio</b>	Cundinamarqués/a
<b>Capital</b>	Bogotá
<b>Límites</b>	
<b>Norte</b>	Boyacá
<b>Sur</b>	Meta y Huila
<b>Este</b>	Casanare
<b>Oeste</b>	Tolima y Caldas

**Fuente:** Elaboración Propia a partir de la información publicada en el Sitio Oficial del departamento.

**Tabla 2 - Información General**

<b>Santander</b>	
<b>Superficie</b>	30.537 Km <sup>2</sup>
<b>Gentilicio</b>	Santandereana/o
<b>Capital</b>	Bucaramanga
<b>Límites</b>	
<b>Norte</b>	Bolívar y Cesar
<b>Sur</b>	Boyacá
<b>Este</b>	Norte de Santander
<b>Oeste</b>	Antioquia

**Fuente:** Elaboración Propia a partir de la información publicada en el Sitio Oficial del departamento.

**Tabla 3 - Información General**

<b>Boyacá</b>	
<b>Superficie</b>	23.189 Km <sup>2</sup>
<b>Gentilicio</b>	Boyacense
<b>Capital</b>	Tunja
<b>Límites</b>	
<b>Norte</b>	Santander y Norte de Santander,
<b>Sur</b>	Cundinamarca
<b>Este</b>	Casanare
<b>Oeste</b>	Caldas y Antioquia.

**Fuente:** Elaboración Propia a partir de la información publicada en el Sitio Oficial del departamento.

**Tabla 4- Información General**

<b>Arauca</b>	
<b>Superficie</b>	23.818 Km <sup>2</sup>
<b>Gentilicio</b>	Araucana/o
<b>Capital</b>	Arauca
<b>Límites</b>	
<b>Norte</b>	Venezuela
<b>Sur</b>	Vichada y Casanare,
<b>Este</b>	Venezuela
<b>Oeste</b>	Boyacá

**Fuente:** Elaboración Propia a partir de la información publicada en el Sitio Oficial del departamento.

**Tabla 5- Información General**

<b>Casanare</b>	
<b>Superficie</b>	44.640 Km <sup>2</sup>
<b>Gentilicio</b>	Casanareña/o
<b>Capital</b>	Yopal
<b>Límites</b>	
<b>Norte</b>	Arauca
<b>Sur</b>	Meta y Vichada
<b>Este</b>	Boyacá
<b>Oeste</b>	Cundinamarca

**Fuente:** Elaboración Propia a partir de la información publicada en el Sitio Oficial del departamento.

## 2.2 SATÉLITE LANDSAT 7 ETM +

Satélite diseñado para observar la tierra con alta precisión, fue puesto en órbita el 15 de abril de 1999 por Estados Unidos con un sensor ETM+ (Enhanced Thematic Mapper Plus), ver en [11]; considerado como uno de los satélites más estables y con elevados estándares de calibración lo que le permitió generar datos de excelente calidad y de fácil acceso al público en general. Sus principales características se indican en la Tabla 6.

**Tabla 6 - Información General Satélite Landsat 7**

Atributo	Descripción
Resolución espacial	30 metros
Orbita	705 +/- 5 km (en el ecuador) sol-síncrono
Inclinación de la órbita	98,2 +/- 0,15
Período de órbita	98,9 minutos
Ciclo de repetición de pista a tierra	16 días (233 órbitas)
Tipo de Producto	Imagen L1T (Precisión, Terrain Corrected) - ortorrectificada
Formato de Imagen	Archivo GeoTIFF
	Metadato en formato texto
Espaciamiento de pixel	15 m Pancromático
	30 m Multiespectral
Tamaño de Producto	185 x 185 km
Bandas / Longitud de Onda	Azul (Blue) / 450 – 50 nm
	Verde (Green) / 530 – 590 nm
	Rojo (Red) / 630 – 690 nm
	Infrarrojos / 700 – 900 nm
	SWIR 1 / 1550 –1750 nm
	Infrarrojos térmicos / 10400 – 12500 nm
	SWIR 2 / 2090 –2350 nm
	Pancromático (Pan) / 520 – 900 nm
Resolución Radiométrica	12 bits

**Fuente:** Elaboración Propia a partir de los aspectos técnicos de las imágenes Landsat publicados por el INEGI.

## 2.3 IMAGEN SELECCIONADA

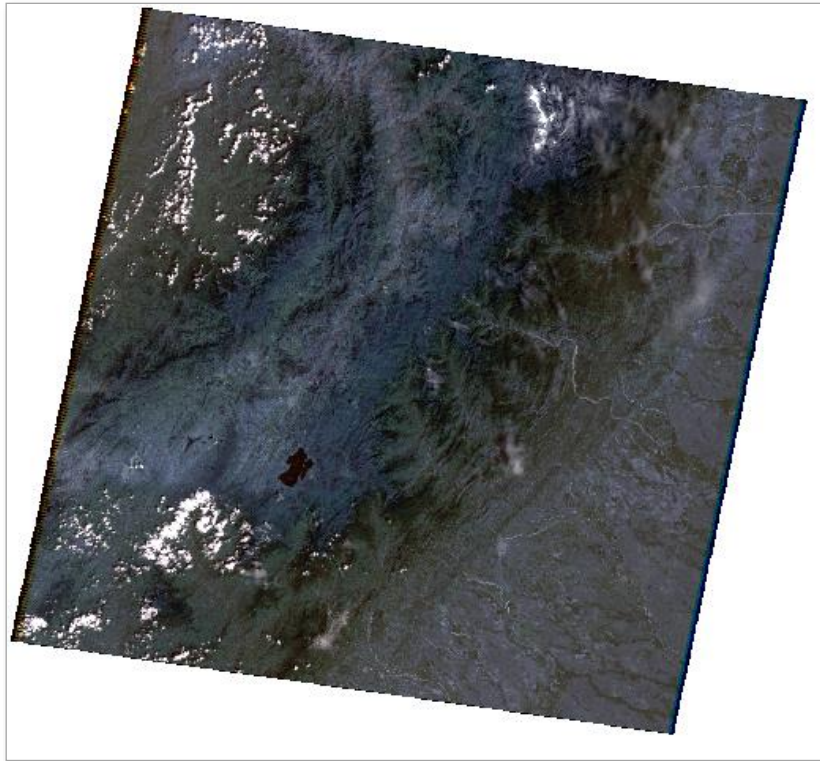
En la Tabla 7 se presentan las principales características de la imagen Landsat 7 seleccionada.

**Tabla 7 - Principales Datos**

Atributo del conjunto de datos	Valor de atributo
Identificador del producto Landsat	LE07_L1TP_007056_20030104_20170127_01_T1
Identificador de escena Landsat	LE70070562003004AGS00
Fecha de Adquisición	4/01/2003
Corrector de línea de exploración	ON
Categoría de la colección	T1
Número de colección	1
Modo Sensor	N/A
Camino WRS	7
WRS Row	56
Fecha L-1 Generado	2017/01/27

Atributo del conjunto de datos	Valor de atributo
Hora de inicio	2003:004:14:54:16.1086875
Hora de finalización	2003:004:14:54:43.2061874
Identificador de estación	AGS
Indicador Día / Noche	DAY

Fuente: UGSS – Science for a changing world.



**Fig. 2.** Localización general zona de estudio.

Fuente: UGSS – Science for a changing world.

## 2.4 COBERTURAS DE LA TIERRA

Las coberturas de la tierra son regiones que se pueden acotar después de elaborar un análisis de respuesta espectral, con el fin de agrupar aquellas áreas que poseen las mismas características diferenciando una de otra [12].

La metodología Corine Land Cover (CLC), realiza un inventario de las coberturas de la tierra, comparando las características físicas para elaborar de mapas a distintas escalas; es así como surge la necesidad de desarrollar e implementar software especializado para obtener productos que cumplan la normatividad vigente.



## **2.5 CLASIFICACIÓN NO SUPERVISADA**

Método que realiza una búsqueda automática agrupando valores uniformes dentro de una imagen; es decir, a partir de los niveles digitales crea varios “cluster” con píxeles que tienen un comportamiento espectral similar [15]. Es importante señalar que el analista debe indicar cuál es el significado temático de las clases espectrales generadas, puesto que el programa no lo detecta y puede establecer categorías de poco interés para el proyecto. Entre los algoritmos más conocidos se destacan ISODATA y K-means.

## **2.6 CLASIFICACIÓN SUPERVISADA**

Este método es bastante útil cuando se posee un conocimiento previo de la zona que comprende la imagen a trabajar, ya que se definen manualmente las áreas de entrenamiento que deben ser significativas con la cantidad de categorías que se pretenden generar [16]. Aquellas áreas tienen que ser correctamente identificadas y cubrir todas las clases, ser homogéneas, altamente correlacionadas y se realizan análisis estadísticos unibanda para verificar y controlar que los valores se asemejen a una distribución normal. Entre los algoritmos clásicos se destacan: Clasificador por mínima distancia, por paralelepípedos y por máxima probabilidad

## **2.7 R**

Es un entorno de software libre orientado a objetos implementando en el lenguaje S creado en los 90, es un proyecto colaborativo y abierto dirigido al análisis estadístico que permite almacenar y manipular datos con gran cantidad de herramientas y gráficos [13]. Hoy en día, se podría decir sin lugar a duda que R es el lenguaje más usado por los estadísticos mundiales, su versatilidad y disponibilidad para cualquier sistema operativo hace que sea una alternativa para procesar grandes cantidades de datos con óptimos resultados; su manejo casi intuitivo, hace que el usuario no requiera mucha experticia programando, por lo que muchas personas de diferentes áreas del conocimiento recurren a este programa y se diversifica en múltiples campos [7] como la percepción remota.

## **2.8 PCI**

Es un software especializado en el procesamiento digital de imágenes (PDI) Satelitales que permite clasificar imágenes multiespectrales, contiene funciones analíticas de geoprocésamiento raster y un entorno de programación visual [14]. Se adapta a las necesidades y requerimientos específicos del usuario, aunque para su manipulación se requiere conocimientos previos en el programa; permite realizar visualización,

realce, corrección y mapeo de imágenes, genera ortofotografías, crea mosaicos, elabora análisis en teledetección y modelamiento además de contener herramientas de radar. Genera resultados de alta calidad posicionándolo en el mercado como una de las aplicaciones con alta capacidad procesamiento geoespacial.

## **2.9 TIPOS DE INVESTIGACIÓN**

Este proyecto se enmarca en los siguientes tipos de investigación:

- Descriptiva: El proyecto se orienta inicialmente a entender cómo se realizan algoritmos de clasificación y obtener capacitación en los programas seleccionados.
- Bibliográfica: Se sustenta la base teórica de la investigación, mediante consultas a fuentes bibliográficas, textos, revistas, documentos, así como también fuentes informáticas e Internet.

## **2.10 ACTIVIDADES**

La metodología que se implementó en este proyecto se divide en 7 actividades, presentadas a continuación:

- a) Selección del sensor satelital con el que se va a trabajar, este se debe ajustar a las necesidades del proyecto.
- b) Selección de la imagen satelital, la cual debe que contener variedad de coberturas para mostrar la capacidad de procesamiento de los programas seleccionados.
- c) Implementación de algoritmos de clasificación para obtener datos cualitativos y cuantitativos de la siguiente forma:

Programa R:

- Cargar librerías y datos.
- Obtener de estadísticas.
- Combinación de bandas.
- Clasificación no supervisada por el método de k-means.
- Clasificación supervisada por el método de RandomForest.

Programa PCI:

- Estadísticas unibanda.

- Datos generales de la imagen Multiespectral.
  - Realización del proceso de corrección atmosférica de la imagen, para eliminar los ruidos y/o distorsiones causados por los efectos atmosféricos y mejorar la calidad visual de imagen. Se obtienen dos correcciones, la primera por radiometría a partir de los datos de calibración del sensor y la segunda por reflectancia a partir de la estimación de los valores de cada pixel en función del análisis de los valores de irradiancia solar, distancia y fecha de captura de la imagen.
  - Combinación de bandas.
  - Clasificación no supervisada.
  - Clasificación supervisada.
- d) Calificación de los programas empleados.
- e) Comparación de los resultados con el fin de describir las ventajas y desventajas encontradas.
- f) Generación de mapa temático.
- g) Conclusiones.

### **3. RESULTADOS**

Este proyecto no intenta determinar cuál de los programas propuestos, resulta ser el mejor al momento de generar clasificaciones no supervisadas y supervisadas, por el contrario, pretende mostrar las ventajas y desventajas al trabajar con el uno o el otro. A continuación, se muestra el proceso realizado con cada programa y se presenta una tabla de calificación con los aspectos más relevantes que a juicio personal le permitirán a cualquier usuario seleccionar el software que más se ajusta a sus necesidades.

#### **3.1 PROGRAMA R**

- a) Cargue de librerías y datos.
- En este programa, es necesario determinar que paquetes de librerías sirven para llevar a cabo las clasificaciones propuestas, cada paquete ofrece una colección de códigos y funciones que facilitan el acceso en el entorno de desarrollo integrado para programar el proyecto; por tal razón se descargaron y utilizaron librerías especializadas en datos espaciales y se enlistan a continuación:

- Library Raster: Implementa funciones básicas y de alto nivel para procesar archivos grandes.
- Library rgdal: Activa operaciones de transformación y proyección de datos.
- Library randomForest: Usa el método de bosques aleatorios de clasificación, supervivencia y regresión; logrando obtener un modelo de predicción para variables categóricas y un modelo de regresión para variables numéricas.
- Library RColorBrewer: Proporciona esquemas de color para mapas.

Dichas librerías permitieron cargar, manipular, visualizar y modelar la imagen Landsat 7, usando líneas de código apropiadas para obtener los resultados esperados.

b) Estadísticas.

Se inicio el proceso, programado líneas de código que permitieron cargar la imagen en el software y se determinó que era necesario utilizar un factor de conversión de los datos para facilitar el almacenamiento de los archivos, esto agilizo el despliegue de información. Paso seguido se realiza la exploración de los datos con histogramas y un diagrama de dispersión para establecer las relaciones espaciales existentes entre las bandas en función de una muestra aleatoria de pixeles; el tamaño de la muestra se definió con la resolución espacial y con la calidad de información aportada por cada banda, de esta manera se define que el tipo de correlación es directa entre las bandas y se establece como favorable continuar con los análisis. Para facilitar el proceso, los datos espaciales fueron asignados en una variable y se obtiene la información básica para cada banda. Ver Fig. 3.

```
> rbrick
class       : RasterStack
dimensions  : 6971, 7801, 54380771, 8  (nrow, ncol, ncell, nlayers)
resolution  : 30, 30  (x, y)
extent      : 646485, 880515, 535785, 744915  (xmin, xmax, ymin, ymax)
coord. ref. : +proj=utm +zone=18 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0
names       : B3.1, B2, B3.2, B4, B5, B6_VCID_1, B6_VCID_2, B7
min values  : 0, 0, 0, 0, 0, 0, 0, 0
max values  : 255, 255, 255, 255, 255, 255, 255, 255
```

**Fig. 3.** Información básica por bandas.

**Fuente:** Elaboración Propia a partir del proceso en R 2017.

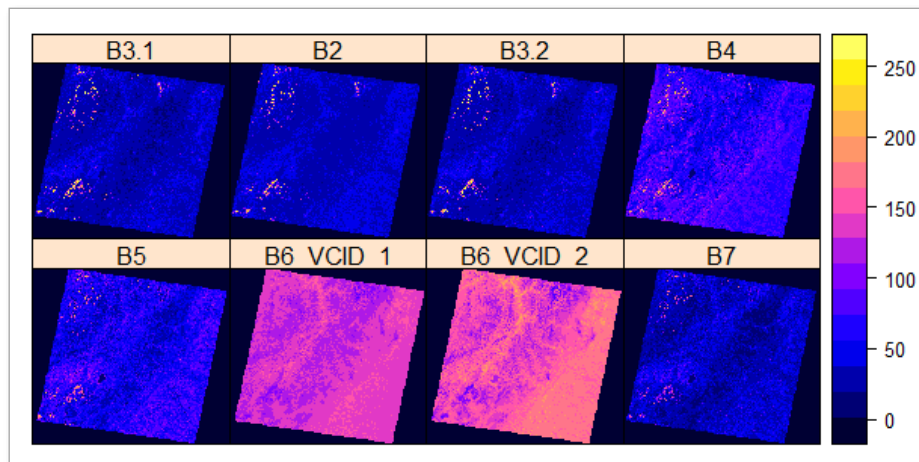
Se calculan las estadísticas como se muestra en la Fig. 4, para describir las principales características y comprender el comportamiento espacial de los datos con el propósito de facilitar su aplicación en las clasificaciones propuestas, de acuerdo con la información aportada por los cuartiles se puede inferir que distribución de frecuencias para los niveles digitales en cada banda en términos generales tienen un buen contraste visual, son pocas la señales no deseadas y es escasa la presencia de nubes. De igual forma, se especializan los datos para visualizarlos y estimar la distribución de cada banda en la zona de estudio. Ver Fig. 5. Se pueden apreciar con tonalidades oscuras los niveles digitales cercanos a cero, con tonos rosados los niveles digitales cercanos a 125 y con tonalidades amarillas los niveles digitales cercanos a 250.

```
> summary(rbrick)
```

	B3.1	B2	B3.2	B4	B5	B6_VCID_1	B6_VCID_2	B7
Min.	0	0	0	0	0	0	0	0
1st Qu.	0	0	0	0	0	0	0	0
Median	25	30	25	60	46	127	142	24
3rd Qu.	34	37	34	76	65	138	163	37
Max.	255	255	255	255	255	171	221	201
NA's	0	0	0	0	0	0	0	0

**Fig. 4.** Estadísticas Unibanda.

**Fuente:** Elaboración Propia a partir del proceso en R 2017.



**Fig. 5.** Espacialización de las bandas.

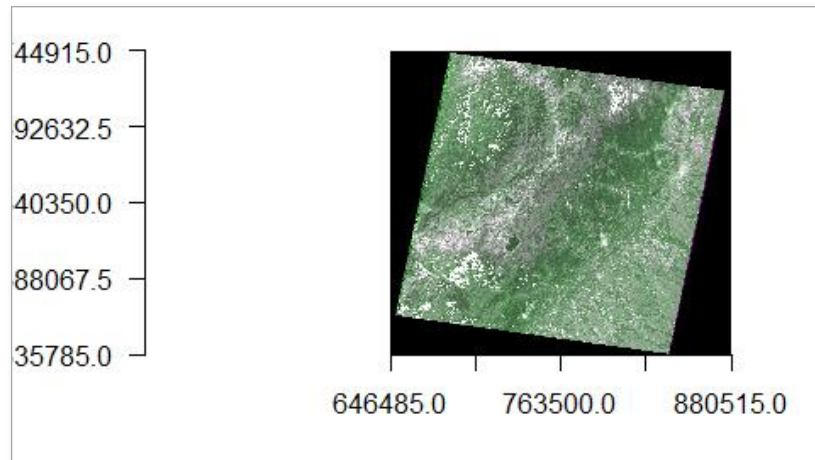
**Fuente:** Elaboración Propia a partir del proceso en R 2017.

#### c) Combinación de bandas.

Para visualizar los cambios presentes en una imagen satelital, se realizan combinaciones entre las bandas multiespectrales, lo que permite discriminar elementos de una forma rápida, obteniendo información de la superficie

terrestre en función de su espectro de emisión. Estas imágenes se crean definiendo el paso de las bandas en uno de los tres cañones de color rojo, verde y azul; así las longitudes de onda tomaran diferentes tonos de pigmentación y se pueden interpretar las coberturas vegetales, usos del suelo, localización de cuerpos de agua, entre otros.

- Combinación color natural RGB 321



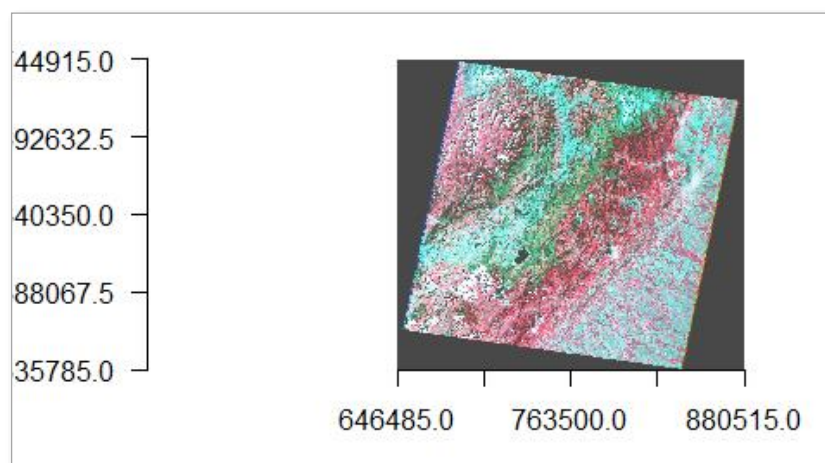
**Fig. 6.**Combinación color natural RGB 321.

**Fuente:** Elaboración Propia a partir del proceso en R 2017.

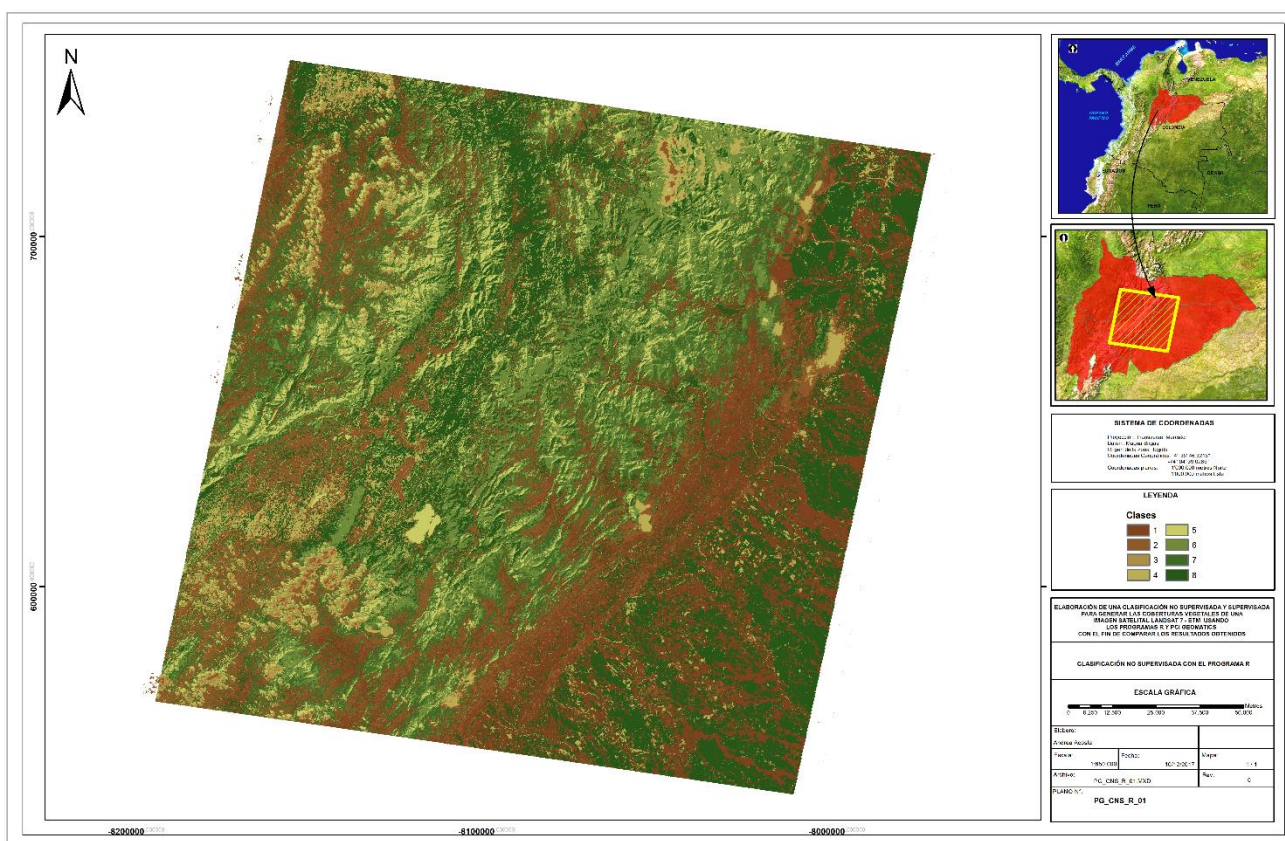
Esta combinación es la que más se aproxima a los colores reales de la superficie terrestre y solo utiliza las bandas que hacen parte de espectro electromagnético visible. Como se observa en la Fig. 6, los tonos azules oscuros señalan aguas pocos profundas y con escasos sedimentos en suspensión, la vegetación se muestra en tonalidades verdes, el suelo aparece en tonos marrones, el blanco indica la presencia de nubes y se pueden identificar a simple vista zonas urbanas.

- Combinación falso color 432

Esta combinación permite visualizar la vegetación y las zonas urbanas, como se muestra en la Fig. 7 el blanco muestra nubes presentes en la imagen, en tonos grises áreas pobladas y suelos desnudos, en color azul oscuro superficies de agua, en tono marrón vegetación arbustiva, en color rosado vegetación en estado de crecimiento y pastos, y en tonos magenta y rojo vegetación vigorosa.



**Fig. 7.** Combinación color natural RGB 432.  
**Fuente:** Elaboración Propia a partir del proceso en R 2017.

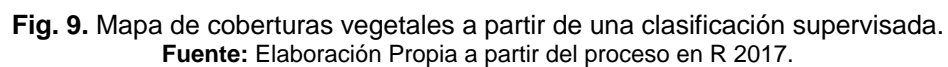


**Fig. 8.** Mapa de coberturas vegetales a partir de una clasificación no supervisada.  
**Fuente:** Elaboración Propia a partir del proceso en R 2017.



- Se trabajó con el método k-means por su eficiencia para manejar grandes conjuntos de datos espaciales; el algoritmo agrupó píxeles en cluster con características espaciales similares, definió centroides y midió la distancia euclidiana entre ellos generando particiones y creando nuevos cluster; este proceso se llevó a cabo iterativamente hasta que no existieron nuevas reasignaciones ni variaciones. Para la Fig. 8 se definieron 8 clases.

- Este estudio nunca planteo realizar un trabajo de campo, ni obtener información cartográfica del área de estudio, por tal razón se tomó como fuente primaria la clasificación no supervisada obtenida anteriormente, identificando zonas con coberturas vegetales homogéneas para ser usadas como polígonos de entrenamiento en el proceso de clasificación.

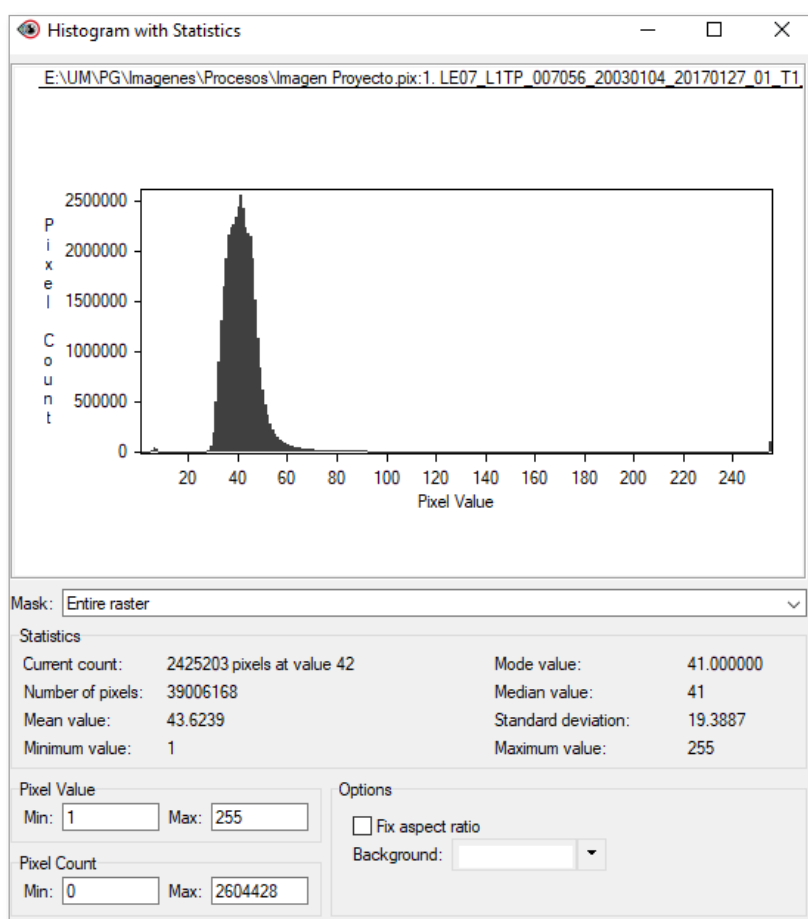




Se utilizó el algoritmo de Random Forest por su capacidad de organizar grandes cantidades de datos espaciales con exactitud; empleando una combinación de árboles predictores, en la que cada árbol dependió del valor del píxel y de un polígono con propiedades espectrales similares, el algoritmo mejora la precisión en la clasificación al integrar aleatoriedad en la construcción de cada clasificador individual. En la Fig. 9 se definieron 11 clases.

### 3.2 PROGRAMA PCI

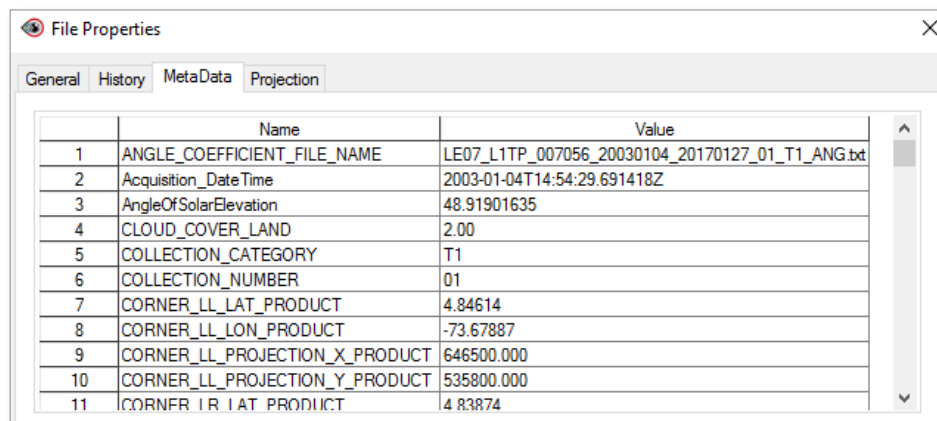
#### a) Estadísticas Unibanda



**Fig. 10.** Estadísticas Unibanda – Banda 1  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

En PCI se cargaron las bandas pancromática y multiespectrales que componen la imagen Landsat 7, con el fin de visualizar las estadísticas y los histogramas para cada banda como se muestra en la Fig. 10. A partir de estos datos se puede inferir de manera general que las desviaciones estandar presentan dispersiones bajas, es decir se tiene poca riqueza visual y es por esta razón las imágenes tienden a verse oscuras al estar concentradas en los primeros niveles digitales; los histogramas son asimétricos positivos y multimodales; gráficamente se determina que la curtosis tiene una distribución leptocúrtica.

#### b) Imagen Multiespectral

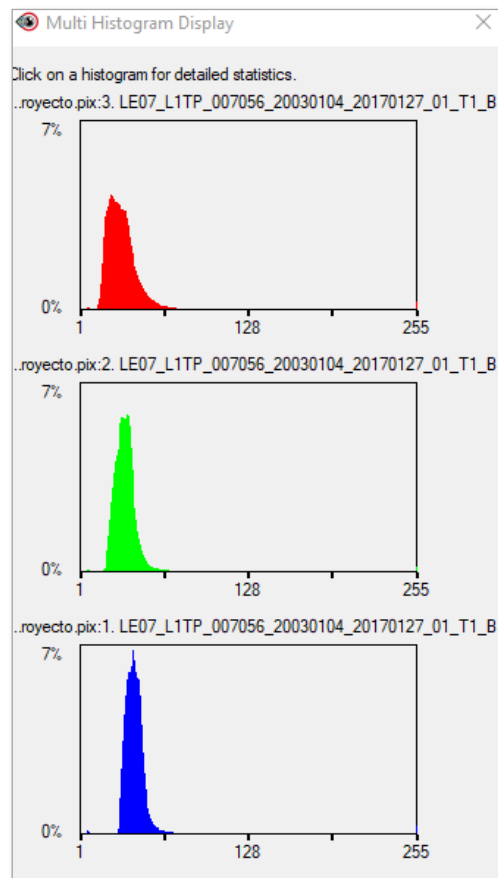


	Name	Value
1	ANGLE_COEFFICIENT_FILE_NAME	LE07_L1TP_007056_20030104_20170127_01_T1_ANG.txt
2	Acquisition_DateTime	2003-01-04T14:54:29.691418Z
3	AngleOfSolarElevation	48.91901635
4	CLOUD_COVER_LAND	2.00
5	COLLECTION_CATEGORY	T1
6	COLLECTION_NUMBER	01
7	CORNER_LL_LAT_PRODUCT	4.84614
8	CORNER_LL_LON_PRODUCT	-73.67887
9	CORNER_LL_PROJECTION_X_PRODUCT	646500.000
10	CORNER_LL_PROJECTION_Y_PRODUCT	535800.000
11	CORNER_UR_LAT_PRODUCT	4.83874

**Fig. 11.** Estadísticas Unibanda – Banda 1  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

Se combinaron las bandas obteniendo una imagen satelital en bruto y se visualizaron los metadatos como se observa en la Fig. 11; de esta manera se pueden obtener información general y datos básicos como fecha de adquisición, nombre de la imagen, datum, efemerides, formato, entre otros.

Los histogramas de la Fig. 12 son estrechos, asimétricos positivos, las medidas de tendencia central no están concentradas en la moda, es multimodal lo que indica que existen varias clases y objetos presentes en la escena, la imagen en bruto se concentra en los primeros niveles digitales y al desplegarla en el programa se ve un poco oscura, por tal razón antes de realizar las clasificaciones propuestas se aplicó un realce adaptativo que permitió diferenciar correctamente las clases presentes, gráficamente la curtosis se clasifica como leptocúrtica es decir que hay alta variación en los niveles digitales y los picos altos muestran zonas sin valor o background.



**Fig. 12.** Histogramas de la imagen multiespectral.  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

Los datos almacenados en una imagen Landsat 7 corresponden a niveles digitales como indica en la Fig.13, estos valores no representan ninguna variable biofísica, por lo tanto si se elaboran analisis con estos datos crudos los resultados pueden estar acompañados de varios errores, ya que no se puede obtener informacion de la reflectancia espectral en la superficie terrestre, visibilidad del suelo y temperatura.

Es importante en la inspección inicial, determinar los errores que están afectando la imagen seleccionada; los errores que por lo general se presentan son fallas en el sensor que generan pixeles errados por tal razón se debe hacer una corrección radiométrica, interferencia de la atmosfera que afecta el valor del pixel y se realiza una corrección atmosférica, por último, el movimiento propio del satélite genera distorsiones en la imagen y se lleva a cabo una corrección geométrica.

**Numeric Values**

☒ Raw data ☐ Enhanced data

...1092001.pix: [3]LE07\_L1TP\_007056\_20030104\_20170127\_01\_T1\_B3-Red

	3905	3906	3907	3908	3909	3910	3911
3483	7.51858	6.759	5.49304	4.48027	3.72069	3.97388	4.73346
3484	7.26539	6.25262	5.23985	4.22707	4.98665	5.74623	5.49304
3485	7.51858	6.25262	5.49304	3.72069	4.98665	5.99942	6.759
3486	7.26539	5.99942	6.25262	4.73346	3.97388	5.23985	5.99942
3487	6.759	5.49304	5.23985	5.23985	3.2143	3.4675	3.72069
3488	6.25262	5.23985	3.97388	4.48027	4.22707	3.72069	3.97388
3489	5.74623	4.73346	3.97388	3.2143	3.4675	5.99942	7.51858

...1092001.pix: [2]LE07\_L1TP\_007056\_20030104\_20170127\_01\_T1\_B2-Green

	3905	3906	3907	3908	3909	3910	3911
3483	8.90516	8.36067	7.54394	6.18271	5.91046	5.91046	6.45495
3484	8.63292	7.81618	6.99945	5.91046	6.7272	7.54394	7.27169
3485	8.08843	6.99945	6.7272	5.36597	6.99945	8.08843	8.08843
3486	7.27169	6.45495	7.54394	6.45495	5.91046	6.99945	7.54394
3487	6.7272	5.91046	6.99945	6.7272	5.36597	5.36597	5.63822
3488	6.7272	5.91046	6.18271	6.18271	5.91046	5.36597	5.63822
3489	6.7272	6.18271	5.36597	4.54924	5.36597	7.54394	9.17741

...1092001.pix: [1]LE07\_L1TP\_007056\_20030104\_20170127\_01\_T1\_B1-Blue

	3905	3906	3907	3908	3909	3910	3911
3483	10.0503	8.81711	7.83055	7.58391	7.33727	7.33727	7.33727
3484	9.31039	8.07719	7.83055	7.09063	8.32383	8.81711	8.32383
3485	8.57047	7.83055	8.07719	6.84399	8.07719	9.31039	8.57047
3486	8.32383	7.58391	9.06375	7.33727	7.58391	7.83055	8.07719
3487	8.32383	7.58391	8.07719	8.32383	6.59735	7.09063	7.09063
3488	8.07719	7.58391	7.33727	7.33727	7.58391	7.09063	7.09063
3489	8.07719	7.58391	7.09063	6.84399	7.09063	9.06375	9.80367

Export Close

**Fig. 13.** Datos de la imagen satelital en bruto  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

c) Correcciones elaboradas a la imagen satelital Landsat 7

La imagen trabajada no tuvo problemas de bandeamiento ni líneas perdidas, se tienen valores para todos los pixeles en el área de estudio.

Por otro lado, la imagen tuvo que ser corregida atmosféricamente con el módulo de ATCOR ejecutado directamente el software PCI, que contiene varias rutinas de trabajo que permiten efectuar diferentes tipos de corrección atmosférica, durante dicho proceso se convirtieron los valores de pixeles en reflectancia física medida encima de la atmósfera con la información propia de la imagen.

En primer lugar, se realizó una modificación radiométrica y se configuraron las bandas para ejecutar las rutinas del módulo, paso seguido se eliminó la neblina y se hizo un enmascaramiento de nubes en la misma escala de los valores iniciales pero corregidos; se estimaron los valores de reflectancia a nivel del suelo corrigiendo efectos del terreno y efectos atmosféricos como vapor de agua y aerosoles. Después, se corrigió la imagen en función de las condiciones de iluminación usando como insumo un DEM (Modelo Digital de terreno) Aster con una resolución espacial de 12.5 metros. Se finaliza la corrección

atmosférica con la temperatura superficial, el programa genera por defecto una imagen con el resultado obtenido. Una vez terminado el módulo se miden nuevamente los valores en la imagen satelital corregida como se indica en la Fig.14 y de esta manera se puede iniciar con las clasificaciones propuestas.

**Numeric Values**

☒ Raw data ☐ Enhanced data

Landsat\_7.pix: [1]focus :Imported from 1 on LE07\_L1TP\_007056\_200301

	3905	3906	3907	3908	3909	3910	3911
3483	47	42	38	37	36	36	36
3484	44	39	38	35	40	42	40
3485	41	38	39	34	39	44	41
3486	40	37	43	36	37	38	39
3487	40	37	39	40	33	35	35
3488	39	37	36	36	37	35	35
3489	39	37	35	34	35	43	46

Landsat\_7.pix: [2]focus :Transferred from 1 on LE07\_L1TP\_007056\_200

	3905	3906	3907	3908	3909	3910	3911
3483	39	37	34	29	28	28	30
3484	38	35	32	28	31	34	33
3485	36	32	31	26	32	36	36
3486	33	30	34	30	28	32	34
3487	31	28	32	31	26	26	27
3488	31	28	29	29	28	26	27
3489	31	29	26	23	26	34	40

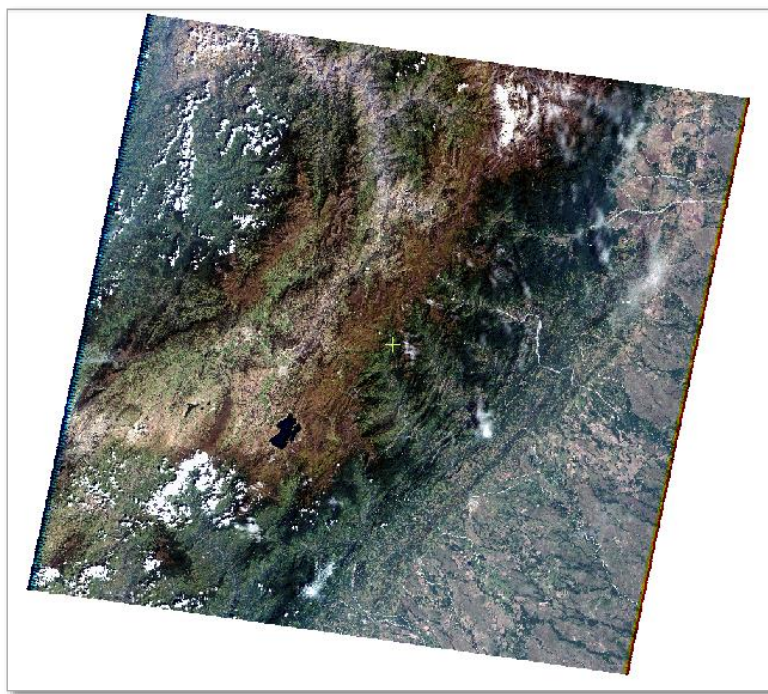
Landsat\_7.pix: [3]focus :Transferred from 1 on LE07\_L1TP\_007056\_200

	3905	3906	3907	3908	3909	3910	3911
3483	36	33	28	24	21	22	25
3484	35	31	27	23	26	29	28
3485	36	31	28	21	26	30	33
3486	35	30	31	25	22	27	30
3487	33	28	27	27	19	20	21
3488	31	27	22	24	23	21	22
3489	29	25	22	19	20	30	36

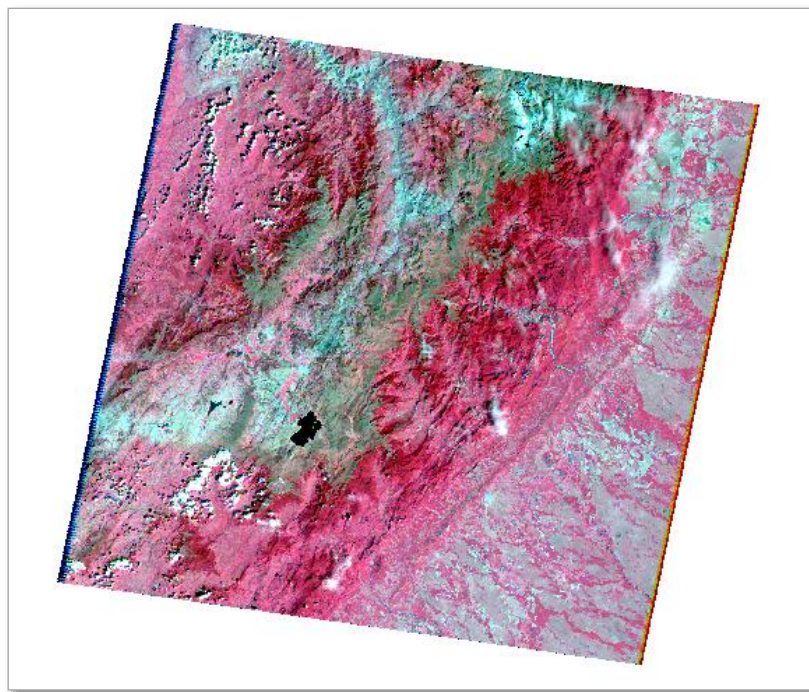
**Fig. 14.** Datos de la imagen satelital Corregida  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

#### d) Combinación de bandas

Al igual que el programa R, se elaboraron las mismas combinaciones para identificar gráficamente los cambios presentes en la imagen satelital; se obtuvieron los mismos resultados llegando a conclusiones similares; la diferencia radica en la calidad de la imagen, el resultado en PCI es de mayor calidad visual, permitiendo diferenciar con un mejor nivel de detalle las clases presentes en la zona de estudio. Ver Fig.15 y Fig.16.



**Fig. 15.**Combinación color natural RGB 321.  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

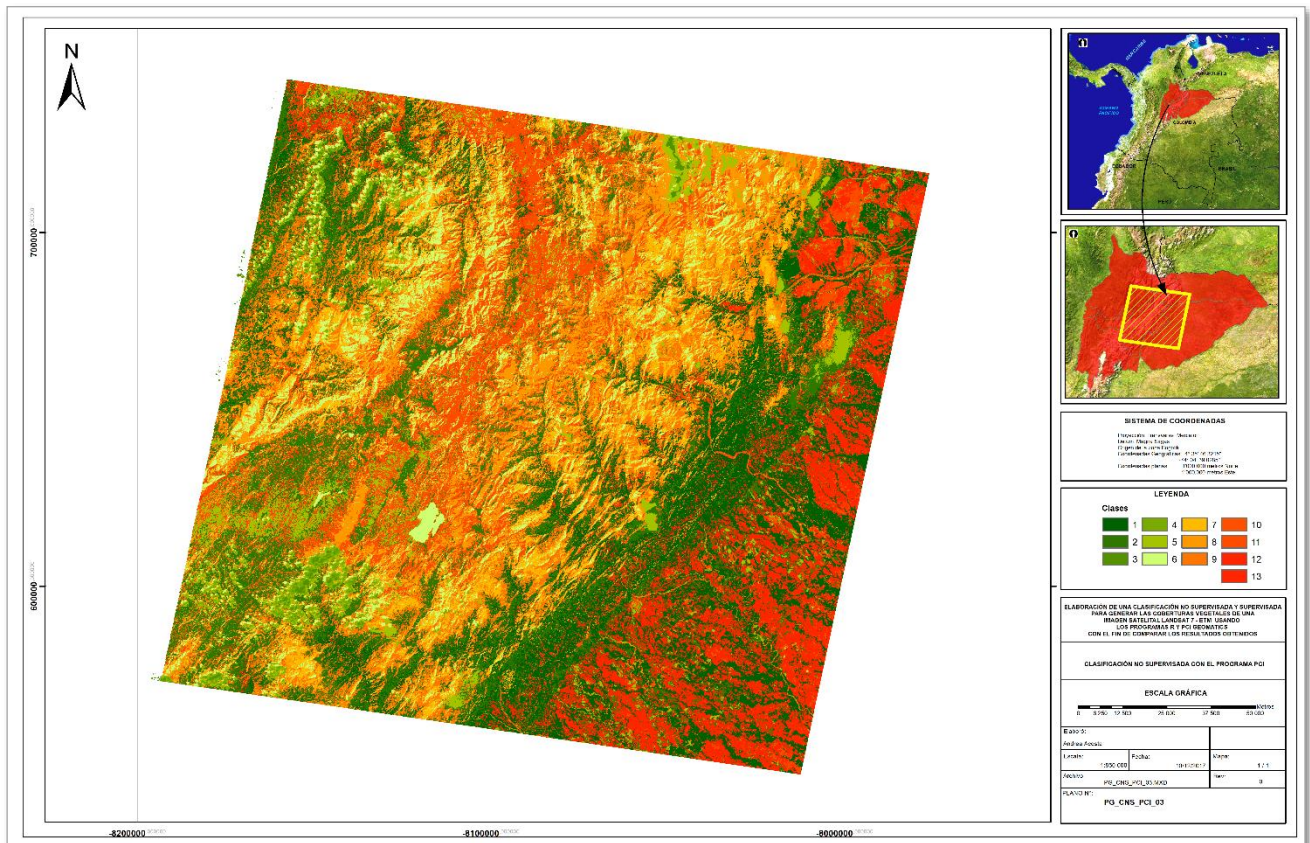


**Fig. 16.**Combinación color natural RGB 432.  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.



e) Clasificación no supervisada

Con la imagen satelital corregida, se procedió a realizar la clasificación no supervisada, el programa indica que se debe crear un nuevo canal para que el resultado sea almacenado sin modificar las características de la imagen multiespectral; se seleccionaron todas las bandas para ser tenidas en cuenta durante el proceso puesto que la imagen fue corregida y hay poca presencia de nubes.



**Fig. 17.** Mapa de coberturas vegetales a partir de una clasificación supervisada.

**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.

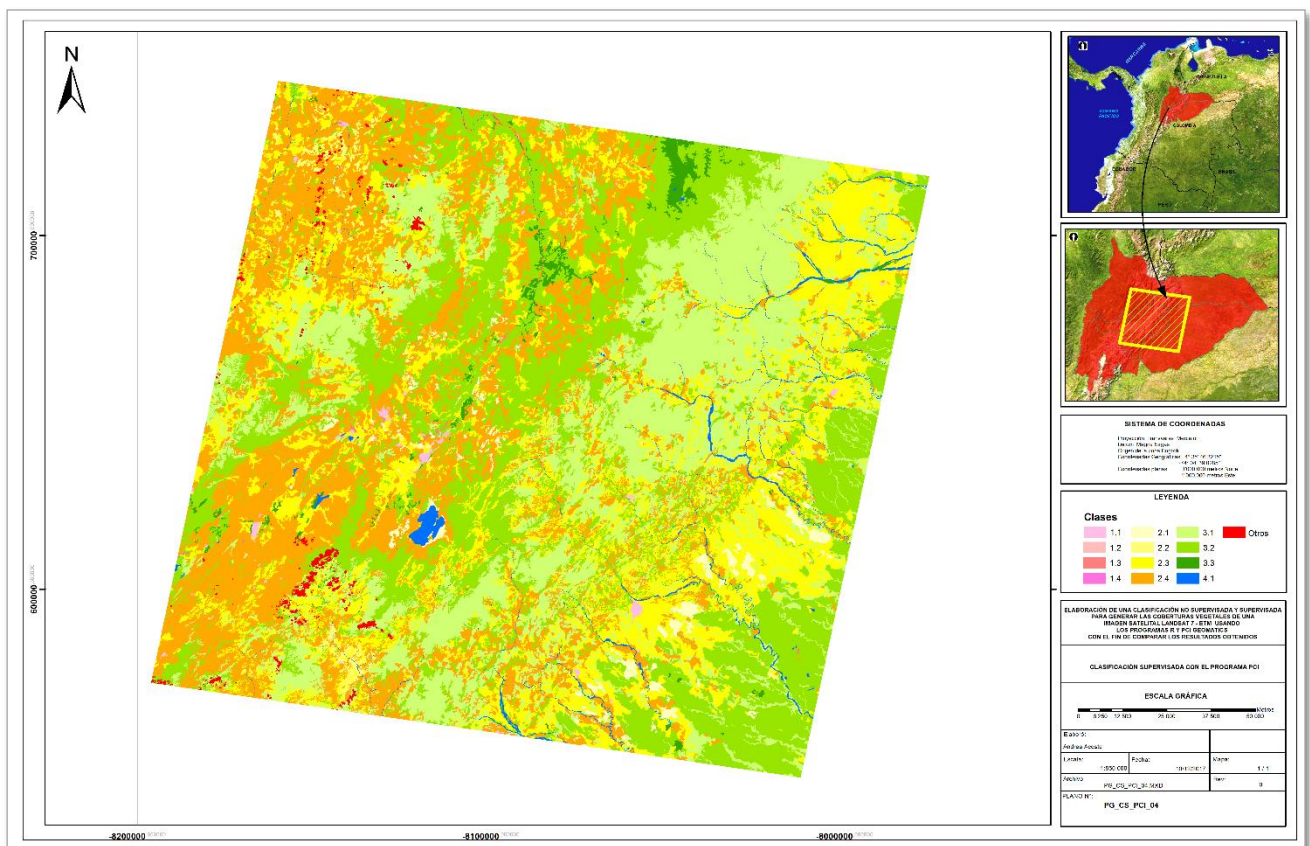
Se decidió no emplear el mismo algoritmo de clasificación para mostrar la versatilidad y capacidad operativa de ambos programas. Por tal razón en PCI se empleó el algoritmo de clasificación Isodata que metodológicamente es similar al algoritmo de k-means pero incorpora la división de zonas con valores dispersos, mezcla áreas cercanas y elimina zonas poco numerosos. PCI solicita definir parámetros para efectuar la clasificación como el número de clases que

se espera encontrar en la imagen para este caso 15, el número de iteraciones, que fueron seleccionadas 20, un valor de tolerancia para la desviación estándar y el valor de background que se dejó en 0. Se crearon 13 clases como se muestra en la Fig.17 y se obtuvieron las firmas espectrales.

En términos generales y acorde con los resultados obtenidos se observa que la imagen presenta zonas heterogéneas que requieren bastante trabajo de edición, por lo que resulta ser útil como aproximación del número de clases que se puede obtener en una clasificación supervisada

f) Clasificación supervisada

La configuración del algoritmo de clasificación supervisado exige crear dos canales el primero permite guardar los datos de training del proceso y el segundo almacena la clasificación final.



**Fig. 18.** Mapa de coberturas vegetales a partir de una clasificación supervisada.  
**Fuente:** Elaboración Propia a partir del proceso en PCI 2017.



Se tomaron muestras de la clasificación no supervisada para afinar los resultados y facilitar el proceso de identificación de las clases presentes en la imagen. El algoritmo seleccionado para este caso fue mínima distancia que en términos generales realiza interpolaciones y análisis de vecindad buscando uniformidad en los polígonos estructurados. PCI pidió que fueran creadas y determinas el número de las clases a identificar para este caso 16, se capturaron varias muestras para cada clase con la herramienta raster seeding, paso seguido se aplicó una máscara para quitar los valores de background presentes en la imagen; estos procesos se repitieron varias veces hasta obtener un resultado satisfactorio dejando finalmente 13 clases como se evidencia en la Fig.18. También se verificó el grado de separabilidad en función de las muestras seleccionadas y se determinó que las coberturas fueron correctamente identificadas, en algunas zonas fue necesario editar polígonos ya que se apreciaban algunos errores. Posteriormente se evaluó la consistencia de los resultados con 30 puntos aleatorios y se generó el respectivo reporte, indicando que solo 6 puntos cayeron en zonas de background, se obtuvo 80% de exactitud en la clasificación que es un porcentaje bastante bueno. Finalmente se suavizaron los pixeles con el algoritmo FMO y se exportó la clasificación a formato shape.

### **3.3 CALIFICACIÓN DE LOS PROGRAMAS EMPLEADOS**

Se hizo una evaluación subjetiva para los programas seleccionados con 7 ítems como se muestra en Tabla 8. Se califica como excelente si cumplió satisfactoriamente, Bueno si cumplió en un término medio y Malo si no cumplió.

A continuación, se define cada ítem propuesto:

- 1) Interfaz de Usuario: Hace referencia a que tan sencillo es para el usuario familiarizarse con el programa.
- 2) Manejo del programa: Indica el nivel de capacitación requerido para manejar el software.
- 3) Formatos de Imagen: Relacionado con la cantidad de formatos que soportan los programas.
- 4) Costo de licencia: Se refiere al tipo de licencia que necesita el software (Licencia paga o libre).
- 5) Tiempos de ejecución: Indica el tiempo que tardan los procesos en finalizar.
- 6) Estadísticas - histogramas: Cantidad de información que muestra cada programa en estas dos temáticas.
- 7) Presentación de resultados: Calidad visual de la imagen final.

**Tabla 8 - Principales Datos**

Ítem	Descripción	Calificación	
		R	PCI
1	Interfaz de Usuario	Excelente	Excelente
2	Manejo programa	Bueno	Bueno
3	Formatos imagen	Bueno	Excelente
4	Costo Licencia	Excelente	Bueno
5	Tiempo de ejecución	Bueno	Excelente
6	Estadísticas - Histogramas	Excelente	Bueno
7	Presentación de resultados	Bueno	Excelente

**Fuente:** Elaboración Propia a partir del proceso en R y PCI 2017.

En términos generales ambos programas cumplieron ampliamente con los resultados esperados, obteniendo buenas calificaciones en los ítems mencionados.

#### **4. CONCLUSIONES**

Se elaboro una calificación subjetiva, para mostrar algunas de las diferencias encontradas durante la elaboración del estudio; el software R aunque no fue diseñado para procesar imágenes satelitales logro ejecutar las clasificaciones propuestas; la interfaz de usuario a diferencia de otros programas usados para programar es amable con el usuario, fácil de manejar y no se requiere de mucha capacitación para entender las rutinas de código y sus funciones; soporta varios formatos de imagen y su licencia es libre; el tiempo de ejecución es un poco lento y algunas actividades de procesamiento no son ejecutas hasta que no se libera memoria en el disco principal del computador, en cuanto a las estadísticas e histogramas es un programa potente y ofrece bastantes alternativas en minería de datos geoespaciales, una limitante es que la imagen resultante no fue de la calidad visual esperada. El software PCI es sencillo de manejar porque viene adaptado por módulos con flujos de trabajo que resultan ser bastante intuitivos si se tiene conocimientos previos en percepción remota, permitiendo detectar errores en las imágenes y solucionarlos durante la ejecución de los procesos, se requiere capacitación en el programa y soporta múltiples formatos de imagen posicionándolo en el mercado por ser capaz de trabajar con varios sensores activos y pasivos; aunque se debe pagar por la licencia considero que es una buena inversión por la cantidad de herramientas que tiene el programa y sus tiempos de ejecución son rápidos; con relación a las estadísticas e histogramas los realiza pero la exploración inicial de datos es algo limitada; por ultimo sus resultados ofrecen una calidad visual excelente.

El proceso y los resultados indican que R es una buena alternativa para explorar los datos estadísticamente hablando, ya que serviría como herramienta para seleccionar y validar con que imagen satelital se debe trabajar, detectando errores antes de realizar procesos que requieren tiempos largos de ejecución. Además, es un programa que está en constante mejora al ser un proyecto colaborativo y abierto. Se podría decir que cuando no se necesite un producto muy elaborado este programa es una buena alternativa para generar coberturas vegetales. PCI es un programa de alto nivel con múltiples herramientas para el procesamiento de imágenes geoespaciales automatizando procesos, útil en teledetección, fotogrametría digital, análisis de imágenes, producción de mapas y generación de mosaicos.

Es importante señalar que el resultado de las clasificaciones depende en gran medida del nivel de experticia del interprete que realice la identificación de las coberturas vegetales, ya que para algunas coberturas tienen niveles de reflectancia bastante similares o firmas espectrales parecidas, así que no basta con tener un software potente de procesamiento.

El usuario finalmente decide con que software trabajar, se sugiere que al momento de tomar una decisión analice que tipo de producto pretende generar, no es lo mismo elaborar un mapa de coberturas vegetales con un alto nivel de detalle que presentar coberturas como información preliminar para definir trabajo en campo o en estudios de poco interés.

## REFERENCIAS

- [1] R. Silva, «Clasificación no supervisada de la cobertura de suelo de la región Sierra de Tabasco mediante imágenes LANDSAT ETM,» *Universidad y ciencia*, vol. 27, nº 1, Abril 2011.
- [2] J. M. Rodríguez, «Análisis multitemporal satelital de los bosques del Carareopón, mediante imágenes Landsat de 1991 y 2002,» *Revistas UDistrital - Colombia forestal*, vol. 9, nº 18, 2005.
- [3] J. L. A. P, «JLAP Bio-Geociencias,» Rutina para el Análisis de Imágenes de Satélite LandSat 5 y LandSat 7 con R Statistical (Landsat package), 2 Marzo 2013. [En línea]. Available: <http://joseluisalvarezpalaciosgeociencias.blogspot.com.co/2013/03/correccion-atmosferica-con-r-statistical.html>.

- [4] J. L. G. Montañez, «RIDUM,» Determinación de las coberturas bosque / no bosque con imágenes ópticas y radar en una zona de Australia, 2017. [En línea]. Available: <http://ridum.umanizales.edu.co:8080/xmlui/handle/6789/3097>.
- [5] R. Lebrete, «Rmixmod: The R Package of the Model-Based Unsupervised, Supervised, and Semi-Supervised Classification Mixmod Library,» *Journal of Statistical Software*, vol. 67, nº 6, Octubre 2015.
- [6] M. A. Gutiérrez, «Clasificación no supervisada de coberturas Vegetales sobre imágenes digitales de sensores Remotos: “landsat – etm,» *Revista Facultad Nacional de Agronomía*, vol. 58, nº 1, 2005.
- [7] S. C. Goslee, «Analyzing Remote Sensing Data in R: The landsat Package,» *Journal of Statistical Software*, vol. 43, nº 4, Junio 2011.
- [8] M. Cruz, «Clasificación no supervisada con imágenes a color de cobertura terrestre,» *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, vol. 44, nº 6, Agosto 2010.
- [9] D. Ballari, «RPubs (brought to you by RStudio),» Analisis y clasificación de imágenes satelitales, Noviembre 2015. [En línea]. Available: [https://rstudio-pubs-static.s3.amazonaws.com/124749\\_390367e0bfc64f1e9c61470ab3e4fff8.html](https://rstudio-pubs-static.s3.amazonaws.com/124749_390367e0bfc64f1e9c61470ab3e4fff8.html).
- [10] UC DAVIS, «R, Introduction to remote sensing data analysis using,» Geospatial and Farming Systems Research Consortium, 2016. [En línea]. Available: [http://gfc.ucdavis.edu/events/arusha2016/\\_static/labs/day4/day4\\_lab1\\_remote-sensing.pdf](http://gfc.ucdavis.edu/events/arusha2016/_static/labs/day4/day4_lab1_remote-sensing.pdf).
- [11] INEGI, «Instituto Nacional de Estadística y Geografía,» Aspectos técnicos de imágenes Landsat, 2015. [En línea]. Available: [http://www.inegi.org.mx/geo/contenidos/imgpercepcion/imgsatelite/doc/aspectos\\_tecnicos\\_de\\_imagenes\\_landsat.pdf](http://www.inegi.org.mx/geo/contenidos/imgpercepcion/imgsatelite/doc/aspectos_tecnicos_de_imagenes_landsat.pdf).
- [12] IDEAM, «Instituto de Hidrología, Meteorología y Estudios Ambientales,» Metodología Corine Land Cover, 2017. [En línea]. Available: <http://www.ideam.gov.co/web/ecosistemas/metodologia-corine-land-cover>.
- [13] J. Contreras, *Introducción a la programación estadística con R para profesores*, 2014.

- [14] GeoSoluciones, «Soluciones Integrales en Geomática - Ingeniería Geoespacial,» PCI Geomatica, 2017. [En línea]. Available: <http://www.geosoluciones.cl/pci-geomatica/>.
- [15] TELEDET, «TELEDET- Percepción Remota,» Imágenes satelitales - clasificación no supervisada, 2017. [En línea]. Available: <http://www.teledet.com.uy/tutorial-imagenes-satelitales/clasificacion-no-supervisada.htm>.
- [16] TELEDET, «TELEDET- Percepción Remota,» Imágenes satelitales - clasificación supervisada, 2017. [En línea]. Available: <http://www.teledet.com.uy/tutorial-imagenes-satelitales/clasificacion-supervisada.htm>.